7-2009

Modeling NO Emissions of an Off-road Diesel Engine Based on Emission Tests

David R. Bostic
Southern Illinois University Carbondale

Dennis G. Watson
Southern Illinois University Carbondale, dwatson@siu.edu

Tony V. Harrison
JPT Integrated Solutions

Modeling NO Emissions of an Off-road Diesel Engine Based on Emission Tests

David R. Bostic[1], Dennis G. Watson[2], and Tony V. Harrison[3]
[1,2]Department of Plant, Soil and Agricultural Systems, MC 4415, Southern Illinois University, Carbondale, IL 62901, USA
[3]JPT Integrated Solutions, Inc, Gainesville, FL 32609, USA
[2]E-mail dwatson@siu.edu

ABSTRACT

Emissions values determined by the ISO 8178 emission certification tests do not necessarily represent emissions of a tractor in operation (Hansson et al., 2001). Rather than using ISO 8178 tests solely for certification, data collected during the tests may be suitable for predicting nitrogen oxide (NO) emissions of an engine operating at constant loads and speeds. Linear multiple regression (LMR) and nonlinear polynomial network (NPN) models were developed with data collected from ISO 8178-4 (1996) test cycle B-type tests (ISO) and an expanded set of tests (EXP) to predict NO emissions from a diesel engine. LMR using the ISO training data (R^2 = 0.94) resulted in over-training of the model, as applied to the evaluation data (R^2 = 0.51). LMR based on the expanded data (R^2 = 0.60) was a better LMR model, when applied to the evaluation data (R^2 = 0.73). NPN using the ISO training data (R^2 = 0.99) resulted in a considerable improvement over the LMR models for the evaluation data (R^2 = 0.81). NPN using the EXP training data (R^2 = 0.96) resulted in the best model when applied to the evaluation data (R^2 = 0.95). When applied to the evaluation data, the mean absolute error of the NPN EXP based model was significantly less than from the NPN ISO based model. The NPN model based on EXP data is recommended for predicting NO. Results from this study suggest data could be collected during ISO 8178-4 emission tests that included additional test modes and modeled with NPN to predict NO emissions for a diesel engine operating at various constant speeds and loads.

Keywords: Diesel emissions, off-road, NO, polynomial network, modeling

1. INTRODUCTION

Non-road, diesel power units are commonly used in agriculture for remote power, such as irrigation pumps. Engine manufacturers are generally required to have engine emissions certified per acceptable standards, such as ISO 8178-4 (1996), and submitted to a governing agency. The goal of ISO 8178-4 was to minimize test modes while ensuring test cycles were representative of actual engine operation (ISO, 1996).

Hansson et al. (2001) compared emissions calculated from use of a 70 kW tractor to ISO 8178-4 emissions values and found ISO 8178-4 test cycle B under estimated carbon monoxide and hydrocarbons by as much as 50%, and over estimated nitrogen oxide emissions by as much as 40%. They concluded it was not possible to design one set of emissions factors that produced representative results for all types of tractors and work operations (Hansson et al., 2001).

Manufacturers typically certify engines intended for diverse applications by using ISO 8178-4 test cycle C or universal test cycle B. Test cycle B specifies 11 test modes for emissions measurement, specifically: 10%, 25%, 50%, 75%, and 100% torque at rated speed and an intermediate speed and no load at low idle. Overall emission values for test cycle B are determined by averaging (other test cycles required weighting) the emissions of the test modes. The resulting emission values would probably not be reliable measures for selecting among engines to power an irrigation pump, as emissions at the actual operating speed and load may vary widely from the certified emission values. Data from individual test modes may be more useful, if the engines in question were certified at a speed and load that matched the actual operating condition. In practice, some engines have been sized to an irrigation pump to operate at rated continuous power, but in many cases, engines operate at less than rated power and may be considerably over-powered for an application. In contrast to passenger or heavy-duty vehicles, irrigation engines operate at a constant speed and load.

Besides using the ISO 8178-4 test cycle B test modes to compute a set of overall emissions values, additional engine operating data from each test mode may be useful for developing a mathematical model to predict emissions of an irrigation engine operating at various load and speed combinations. The engine speed, percent torque, and emissions values for each mode are recorded. If other data from an electronically controlled engine's controller area network (CAN), ambient conditions, torque, and exhaust characteristics were available for each mode, it may be possible to develop a model to predict emissions at various speed and load combinations. Additional test modes may be needed to model emissions at speeds different than the two tested speeds or at loads between the tested loads of test cycle B.

Emission models have been developed for other diesel-powered vehicles. Ramamurthy et al. (1998) fit a polynomial curve to nitrogen oxides (NO_x) emissions based on axle power of a heavy-duty diesel vehicle and described the simple correlation as near-linear. Cooper and Andreasson (1999) used non-linear regression to predict NO_x emissions of a diesel powered passenger ferry. Although fuel rate, engine load, ambient air temperature, relative humidity, barometric pressure, charge air temperature, exhaust temperature and oxygen (O_2) concentration in the exhaust were initially measured, the selected model used only oxygen (O_2) concentration in the exhaust and engine power (kW) for an R^2 of 0.961 (Cooper and Andreasson, 1999). Yanowitz et al. (2002) used test data from a heavy-duty vehicle transient test to predict diesel emissions based on engine power and found a good linear correlation between rates of horsepower increase and particulate matter emissions. Krijnsen et al. (2000) used inputs of engine speed, injection pump rack position, charge air pressure and charge air temperature to successfully model NO_x emissions from a diesel engine using an artificial neural net, a split and fit algoritm, and a nonlinear polynomial model; and demonstrated that the NO_x predictions based on these algorithms were more accurate than a linear model.

1.1. Objective

A study was conducted to compare models derived from two data sets and two modeling methods for predicting NO emissions of a diesel engine operating at constant loads and speeds. For modeling purposes, the target range of engine operation was 1500 rpm to rated speed (2500 rpm) with 10% increments of torque starting at 40% up to 100%. The data sets consisted of data obtained from test modes similar to the ISO 8178-4 test cycle B and an expanded set of test modes with additional loads and speeds. The data included engine operating conditions from the engine's CAN plus torque, emissions, and ambient condition sensors. The modeling methods used were linear multiple regression (LMR) and a nonlinear polynomial network (NPN). The results of the study provide comparative data on the relative suitability of the ISO 8178-4 test cycle B-type data and expanded data for predicting NO emissions of a diesel engine running at a constant load and speed.

ISO 8178 compliant emissions measuring equipment and controlled environment were not available for this study. Since this is an initial modeling study and results were not intended to be compared to emission certification data, this limitation was considered acceptable.

1.2. Nonlinear Polynomial Network Modeling

NPN modeling is a non-parametric, self-organization approach in which underlying relationships of variables are automatically discovered by the NPN algorithm. In this context, a network is a function represented by the composition of many functions (Barron and Barron, 1988) (see Figure 1 for example network). NPN is closely related to the group method of data handling (GMDH) algorithm developed in Kiev, and first published by Ivakhnenko (1968). According to Farlow (1984), Ivakhnenko's work was prompted by the requirement of many mathematical models to know details about a system that are generally unknown. A method was needed that relied on objective methods rather than biases of the researchers (Farlow, 1984). Barron et al. (1984) described early polynomial network software development in the United States as based on the GMDH described by Ivakhnenko (1971).

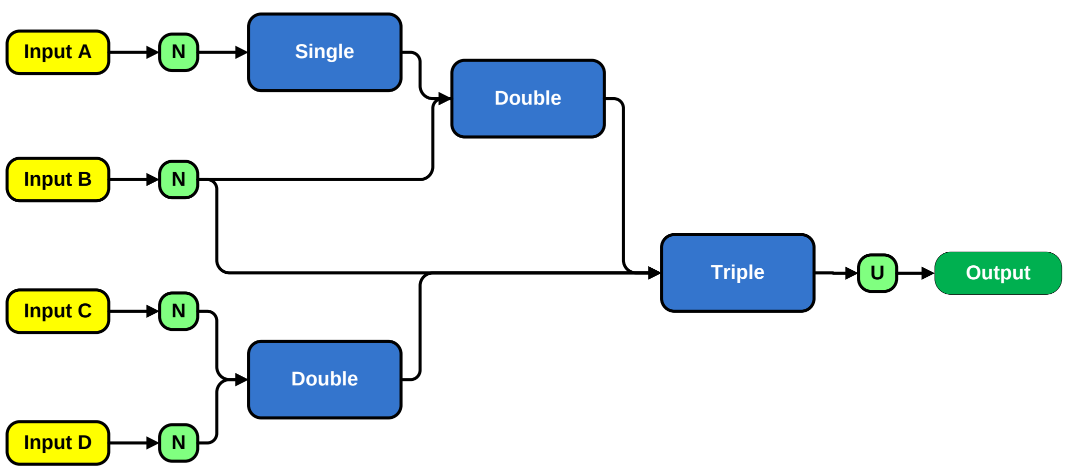

Figure 1. Example of polynomial network with N symbolizing normalizing function, U symbolizing unitizing function, and single, double, and triple indicating the number of inputs in a network node.

NPN software based on GMDH-type algorithms have been described using various terms, including: polynomial network (Barron et al., 1984; Drake et al., 1994; Griffin et al., 1994; Kleinsteuber and Sepehri, 1996), abductory induction (Montgomery, 1989), abductive reasoning networks (Montgomery and Drake, 1991), and abductive polynomial network (Drake and Kim, 1997). More recently, polynomial network software have been classed as data mining tools

(Agarwal, 1999; Kim, 2002; King et al., 1998; and Pyo et al., 2002). Polynomial networks have been used for a wide range of modeling applications, including: defense (Montgomery et al., 1990), financial (Stepanov, 1974; Kim, 2002), medical (Abdel-Aal and Mangoud, 1997; Griffin et al., 1994), process control (Silis and Rozenblit, 1976), and agriculture (Duffy and Franklin, 1975; Ivakhnenko et al., 1977; Lebow et al., 1984; Pachepsky and Rawls, 1999; Reddy and Pachepsky, 2000).

2. EQUIPMENT AND PROCEDURES

2.1. Two emission tests

Two different emission tests were conducted to collect NO and engine operating data for the purpose of comparing the resulting data in modeling NO. The first test was called ISO and was based on ISO 8178-4 (1996) test cycle B. The rated speed was 2500 rpm and an intermediate speed of 1500 rpm was selected. No-load tests were substituted for the 10% torque tests, since testing equipment would not support the low 10% of torque at both speeds. The 11 torque and engine speed combinations were replicated four times for a total of 44 tests. Although ISO 8178-4 does not specify replications, they were added to provide additional data for modeling and model evaluation.

The second emissions test expanded on the ISO test and was called EXP. It was designed to provide more data points by testing loads at 10% intervals from 40% to 100% torque and using additional speed settings. Emissions data were collected while the diesel engine was operated at 0%, 25%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of torque for each engine speed of 1500, 1750, 2000, 2250, and 2500 (rated power) rpm. The 45 torque and engine speed combinations were replicated four times for a total of 180 tests. Other than the torque and speed combinations there were no other differences in equipment and procedures in collecting data for the ISO and EXP emissions tests.

2.2. Equipment

A 2003 John Deere 4045T, 4.5 L, inline four cylinder, EPA Tier 2, turbocharged, electronically controlled diesel engine was used for the emissions tests. Peak torque was 394 Nm at 1400 rpm and rated power was 86 kW (77kW for continuous operation) at 2,500 rpm. This engine was equipped with a SAE J1939 (SAE, 2002) CAN. An Opto 22 SNAP Ultimate I/O programmable automation controller (PAC) was selected for data acquisition. Dearborn Group Technology's Dearborn Protocol Adapter (DPA) model DPA III/i was used to interface between the CAN and the PAC. The DPA was connected to a diagnostic port on the engine wiring harness and to a serial module controlled by the PAC. SAE J1939-71 (SAE, 2002) was referenced to interpret CAN signals and program the PAC to extract engine speed, throttle position, and fuel rate data. The PAC monitored CAN data at 500 ms intervals. Ambient conditions of humidity, temperature and atmospheric pressure were measured with sensors placed within 4 m of the engine and connected to the PAC for data acquisition. This equipment configuration was previously described by Hogan et al. (2007) and Watson et al. (2008).

An M&W P-400B hydraulic dynamometer was used to provide an engine load. A Lebow TMS 9000 torque measurement system was used to measure torque. The system consisted of a rotating torque sensor, mounted between the engine flywheel and the dynamometer driveshaft, and a signal processing module. The output of the signal processing module was connected to the PAC. For no-load tests, the dynamometer driveshaft was disconnected from the torque sensor.

Exhaust emissions were collected and analyzed by an Infrared Industries FGA4000XD gas analyzer (GA). The GA used a chemical cell to measure NO. The GA also measured exhaust temperature, pressure, and air to fuel ratio. Exhaust gasses were collected by connecting a tube to the exhaust system upstream from the muffler. GA output was connected to the PAC. The GA measured NO in parts per million (ppm) and units of g/kWh were calculated as specified by Infrared Industries.

ISO 8178-4 (ISO, 1996) specified the test time for each mode be no less than 10 minutes. This included seven minutes for engine adjustment and stabilization and a minimum of three minutes for data collection. For this study, a minimum of five minutes was used to adjust the engine speed and load, and allow the engine to stabilize. Once the engine stabilized at the desired settings, two additional minutes passed before a data collection period of eight minutes started.

LabVIEW (National Instruments, 2003) was programmed to provide a user interface to the data monitored by the PAC, to collect the data at the specified time intervals, and store the data from each test mode in a comma-separated values file. The LabVIEW program was installed on a computer and used by the operator to control data collection. OPC (object linking and embedding for process control) server software was used to transfer data between the PAC and OPC client of LabVIEW. This system provided real-time data access with updates as frequent as 200 ms. The LabVIEW program automatically recorded CAN, torque, ambient condition, and GA data at 30 second intervals during an eight minute test run. After each test mode ended, the data points were averaged.

SAE standard J1939-71 defined variables potentially available on the CAN (SAE, 2002). Eight variables were available that were related to engine performance. These were included in the 18 variables measured or calculated for the emissions tests (see Table 1).

Table 1. Variables measured and calculated during engine emissions tests and used for modeling.

Variable Name	**Source**
Engine speed (rpm)	CAN*
Percent torque	CAN
Percent load	CAN
Percent friction	CAN
Fuel rate (l/h)	CAN
Engine fuel temperature (deg C)	CAN
Coolant temperature (deg C)	CAN
Intake manifold temperature (deg C)	CAN
Torque (Nm)	Lebow TMS 9000
Power (kW)	Calculated
Ambient temperature (deg C)	Opto 22 ICTD
Relative humidity (%)	Honeywell HIH3610
Barometric pressure (mbar)	Novalynx WS16BP
Exhaust temperature (deg C)	GA
Exhaust pressure (kPa)	GA
Air to fuel ratio	GA
NO (ppm)	GA
NO (g/kWh)	Calculated

* CAN variables may be direct sensor readings or inferred values. Percentage values and fuel rate are typically inferred values based on other data available in the engine controller.

2.3. Modeling procedures

The data from the two emissions tests were combined into three files for modeling. All data from replications one, two, and three of the ISO tests were combined into one file for the ISO training data. Likewise, all data from replications one, two, and three of the EXP tests were combined into one file for the EXP training data. Data from replication four of both the ISO and EXP tests were combined into one file for the evaluation data set. The first 16 variables from Table 1 were used as independent variables (inputs) and NO in g/kWh was the dependent (output) variable. Although the resulting training sample sizes (n = 33 for ISO and n = 135 for EXP) are relatively small for LMR and are expected to result in over fitting, the LMR models are included as a comparison to the NPN models—which have been found to be more efficient than LMR with small sample sizes (Stepanov, 1974).

SAS® 9.1 (SAS, 2007) was used to compute correlation and regression coefficients and significance ($\alpha = 0.05$) for the 16 inputs to NO. SAS® 9.1 (SAS, 2007) model selection methods of highest R^2, highest adjusted R^2, and stepwise regression were used for each of the ISO and EXP training data. Models based on the three selection methods were applied to the evaluation data to predict NO and the best model based on highest R^2 and lowest root mean squared error (RMSE) was selected for each of the ISO and EXP training data.

Two NPN models were developed—one each for the ISO training and EXP training data. ModelQuest® (MarketMiner, 2004) software was used to complete the steps to derive the NPN model. ModelQuest software has been used by other researchers, including Abdel-Aal and Mangoud (1997), Agarwal (1999), Cerullo and Cerullo (2006), Kim (2002), and Reddy and Pachepsky (2000).

The NPN was calculated one layer at a time. The initial (or input) layer consisted of normalizing each of the 16 inputs to a mean of zero and standard deviation of one. For each subsequent layer of the network, each possible combination of inputs from the prior layer was combined into third order polynomial equations with each combination of single, double, and triple inputs, using the following equations (Montgomery, 1989).

$$\text{Single} = w_0 + w_1x_1 + w_2x_1^2 + w_3x_1^3 \quad (1)$$

$$\text{Double} = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2 + w_5x_1x_2 + w_6x_1^3 + w_7x_2^3 \quad (2)$$

$$\text{Triple} = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_1^2 + w_5x_2^2 + w_6x_3^2 + w_7x_1x_2 + w_8x_1x_3 + w_9x_2x_3 + w_{10}x_1x_2x_3 + w_{11}x_1^3 + w_{12}x_2^3 + w_{13}x_3^3 \quad (3)$$

where:
w_i is a coefficient to be determined
x_i is an input variable
Single is an equation with one input variable
Double is an equation with two input variables
Triple is an equation with three input variables

Predicted squared error was used as the selection criterion. Barron (1984) defined PSE as consisting of a squared error term based on the training data and an overfit penalty term as follows.

$$PSE = TSE + 2\sigma_p^2 K/N \quad (4)$$

where:
K is the number of coefficients estimated to minimize TSE
N is the number of training observations
σ_p^2 is the prior estimate of true error variance
TSE is the training squared error
PSE is the predicted squared error

PSE was applied to each of the single, double, and triple equations, along with inputs from the prior layer and original inputs to select the best predictors for input to the next level. The selection criterion was also applied to the resulting network. Until the selection criterion for the network was met, additional layers were added to the network using inputs calculated in the prior layer. As each coefficient was added to reduce the error of the NPN, the over fit penalty increased. The over fit penalty was designed to keep a model from over fitting the training data to the extent that it performs poorly on future observations. Once PSE for a layer increased from the prior layer, the NPN from the prior layer was selected as the best model. The resulting value (normalized NO) was converted to units of g/kWh.

Each of the NPN models was evaluated with the same evaluation data set as the LMR models. The models were compared based on coefficient of determination (R^2), average error, and maximum error. Paired t-tests were also used to determine if absolute error was different between paired models.

3. RESULTS AND DISCUSSION

3.1. Variable correlations

Data from each of the ISO and EXP training data sets were analyzed using correlation. Table 2 lists the input variable names and correlation coefficients for each set of training data. For the ISO training data, correlations for 9 of the 16 input variables with NO were significant at $p < 0.0001$, with 5 additional inputs significant at $p < 0.05$. Only atmospheric pressure and exhaust pressure were not significant. Correlations of the EXP training data found 9 of the 16 input variables significant at $p < 0.0001$, with two additional inputs significant at $p < 0.05$. The five EXP training data inputs that were not significantly correlated to NO included the two from the ISO training data, plus engine speed, percent friction, and relative humidity.

Table 2. Correlation coefficient of each explanatory variable to NO (g/kWh) for each of the ISO and EXP training data sets.

Variable Name	Raw Data Range	Correlation Coefficients (*r*) ISO	Correlation Coefficients (*r*) EXP
Engine speed (rpm)	800 - 2516	-0.51*	-0.07
Percent torque	9 – 90	-0.66**	-0.54**
Percent load	10 – 100	-0.67**	-0.53**
Percent friction	10 – 17	-0.37*	-0.06
Fuel rate (l/h)	0.76 – 26.0	-0.66**	-0.50**
Engine fuel temperature (deg C)	32 – 61	-0.70**	-0.24*
Coolant temperature (deg C)	76 – 86	-0.56*	-0.47**
Intake manifold temp (deg C)	21 – 150	-0.65**	-0.47**
Torque (Nm)	2.9 – 426.5	-0.66**	-0.53**
Power (kW)	0.3 – 93.7	-0.67**	-0.52**
Ambient temperature (deg C)	17 – 35	-0.78**	-0.45**
Relative humidity (%)	9 – 32	0.71**	0.12
Atmospheric pressure (mbar)	29.29 – 29.62	0.05	-0.12
Exhaust temperature (deg C)	27 – 50	-0.47*	-0.27*
Exhaust pressure (kPa)	748 – 759	-0.06	-0.14
Air to fuel ratio	18.4 – 49.9	0.56*	0.41**

* Correlation coefficient was significant ($\alpha = 0.05$) with $p < 0.05$.

** Correlation coefficient was significant ($\alpha = 0.05$) with $p < 0.0001$.

Correlations of each input variable to NO were stronger for the smaller ISO training data set with the exception of atmospheric pressure and exhaust pressure. For the ISO data, the strongest correlations were ambient temperature, relative humidity, and engine fuel temperature with correlations in the absolute value range of 0.70 to 0.78. For the EXP data, the five strongest correlations of -0.50 to -0.54 were for percent torque, percent load, torque, power, and fuel rate.

Strong correlations were expected between pairs of the available inputs such as torque, percent torque, percent load, and power. Table 3 indicates pairs of input variables with an absolute correlation coefficient greater than 0.90 in either the ISO or EXP data set.

Table 3. Pairs of input variables (marked by X) with an absolute correlation coefficient greater than 0.90 in either the ISO or EXP training data set.

	Engine speed	**Percent torque**	**Percent load**	**Percent friction**	**Fuel rate**	**Engine fuel temperature**	**Intake manifold temperature**	**Torque**	**Power**	**Air to fuel ratio**
Engine speed				X		X				
Percent torque			X		X			X	X	X
Percent Load		X			X		X	X	X	X
Percent friction	X									
Fuel rate		X	X				X		X	
Engine fuel temperature	X									
Intake manifold temperature			X		X				X	
Torque		X	X						X	X
Power		X	X		X		X	X		
Air to fuel ratio		X	X					X		

3.2. Linear multiple regression (LMR) models

LMR was used to fit an equation to the combination of the 16 input variables to predict NO. For both the ISO and EXP data sets, the LMR model selection method using the highest R^2 resulted in a higher R^2 and lower RMSE, when applied to the evaluation data, that the models from the highest adjusted R^2 or the stepwise selection methods. For the ISO data set, the resulting LMR equation accounted for approximately 94% of observed variance in NO in the training data ($F_{16,16}$ = 14.68, p < 0.0001, adjusted R^2 = 0.87). Table 4 lists the regression coefficients and standardized coefficients for each input. Inputs of percent load, percent torque, power, intake manifold temperature, engine fuel temperature, and torque had the highest weights, but none of the weights were considered significant at p < 0.05.

Table 4. Regression coefficients and standardized coefficients for each of the ISO and EXP training data sets.

	Regression Coefficients		Standardized Coefficients	
Variable Name	**ISO**	**EXP**	**ISO**	**EXP**
y-intercept	1078.931	404.158*	0	0
Engine speed (rpm)	-0.003	0.005	-0.362	0.433
Percent torque	-0.457	-0.099	-2.642	-0.567
Percent load	0.455	0.191	3.068	1.264
Percent friction	0.957	-1.832	0.612	-0.993
Fuel rate (l/h)	0.134	-2.473	0.213	-3.840
Engine fuel temperature (deg C)	-0.863	-0.260*	-1.196	-0.304
Coolant temperature (deg C)	-0.343	-0.546*	-0.189	-0.240
Intake manifold temp (deg C)	0.173	0.464**	1.455	3.954
Torque (Nm)	-0.039	-0.084	-1.139	-2.538
Power (kW)	-0.285	0.223	-1.730	1.429
Ambient temperature (deg C)	0.446	-0.210	0.334	-0.174
Relative humidity (%)	0.417	-0.245*	0.172	-0.294
Atmospheric pressure (mbar)	-1.474	-0.113	-0.384	-0.075
Exhaust temperature (deg C)	-0.613	-0.267	-0.313	-0.215
Exhaust pressure (kPa)	0.633	-0.265	0.131	-0.174
Air to fuel ratio	-0.324	-0.035	-0.808	-0.094

* Regression coefficient significant ($\alpha = 0.05$) with $p < 0.05$.

** Regression coefficient significant ($\alpha = 0.05$) with $p < 0.0001$.

For the EXP data set, the resulting LMR equation accounted for approximately 60% of observed variance in NO ($F_{16,118} = 11.23$, $p < 0.0001$, adjusted $R^2 = 0.55$). Table 4 lists the regression coefficients and standardized coefficients for each input. Inputs of intake manifold temperature, fuel rate, torque, power, and percent load had the five highest weights. The intake manifold temperature coefficients was significant at $p < 0.0001$. Regression coefficients for engine fuel temperature, coolant temperature, relative humidity, and y intercept were also significant at $p < 0.05$.

The respective regression equations of the ISO and EXP training data were applied to the evaluation data to predict NO. Table 5 summarizes the R^2, mean absolute error, and maximum absolute error of each model applied to the training data and evaluation data.

Table 5. Comparative performance of LMR and NPN models derived from each of ISO and EXP training data sets and the evaluation data.

		Training Data**			Evaluation Data***		
Data Set	**Model Strategy**	**R^2***	**Mean Absolute Error**	**Maximum Absolute Error**	**R^2***	**Mean Absolute Error**	**Maximum Absolute Error**
ISO	Linear Multiple Regression (LMR)	0.94	3.20	11.24	0.51	3.84	19.83
ISO	Nonlinear Polynomial Network (NPN)	0.99	0.47	1.84	0.81	2.79	20.04
EXP	Linear Multiple Regression (LMR)	0.60	1.59	15.13	0.73	3.00	21.12
EXP	Nonlinear Polynomial Network (NPN)	0.96	0.46	6.25	0.95	1.15	8.18

* Coefficient of determination between actual and predicted NO for the respective data set.

** Actual NO (g/kWh) values in the ISO data set ranged from 1.27 to 18.55 with a mean of 4.50 and standard deviation of 4.90. Actual NO values in the EXP data set ranged from 0.01 to 29.17 with a mean of 2.86 and a standard deviation of 3.99.

*** The actual NO (g/kWh) values in the evaluation data set ranged from 1.32 to 35.89 with a mean of 5.34 and a standard deviation of 7.53.

When applied to the evaluation data, the R^2 for the LMR model based on the ISO training data dropped from 0.94 to 0.51. This model was able to explain only half the variation in the evaluation data. Mean absolute error and maximum absolute error increased by multiples of 1.2 and 1.8, respectively, for the evaluation data compared to the training data. Although the ISO training data alone resulted in a strong linear relationship, the model was over-trained and was not as effective at predicting the evaluation data with the additional engine operating conditions. Results of this model indicated ISO 8178-4 test cycle B-type tests modes alone would not be sufficient to model the target range of engine operation.

The R^2 for the EXP based LMR model increased from 0.60 for the training data to 0.73 for the evaluation data. This was the only model to explain more of the variance of NO for the evaluation data than the training data. For the EXP based model, mean and maximum absolute error increased by multiples of 1.9 and 1.4, respectively, when applied to the evaluation data. When comparing mean absolute error of the LMR models applied to the evaluation data, the model based on the EXP data had a significantly lower error ($t_{(55)}$ = 2.736; p = .0084). The additional test modes in the EXP training data resulted in a better model than the ISO-based LMR model, but still left 27% of the variation of NO unexplained.

3.3. Nonlinear polynomial network (NPN) models

The ISO training data were used to develop a NPN model to fit an equation to NO based on the 16 input variables. The resulting NPN is depicted in Figure 2. Of the 16 input variables, only torque, engine fuel temperature, and exhaust temperature were used by the resulting polynomial network to describe the variability in NO. The predicting network accounted for approximately 99% of the observed variance in NO and consisted of the following network of equations:

$$T_n = -1.2661 + 0.0071 * T \tag{5}$$

$$FT_n = -6.7894 + 0.1473 * FT \tag{6}$$

$$EX_n = -17.6296 + 0.4001 * EX \tag{7}$$

$$TR = -0.7817 + 0.5391 * T_n^2 - 0.2811 * T_n^3 - 0.2612 * T_n * FT_n^2 + 0.3318 * EX_n^2 - 0.1478 * EX_n^3 \tag{8}$$

$$NO = 4.4988 + 4.896 * TR \tag{9}$$

where:
T_n is normalized torque
T is observed torque (Nm)
FT_n is normalized engine fuel temperature
FT is observed engine fuel temperature (°C)
EX_n is normalized exhaust temperature
EX is observed exhaust temperature (°C)
TR is the intermediate output of a network node with three inputs
NO is nitrogen oxide emissions (g/kWh).

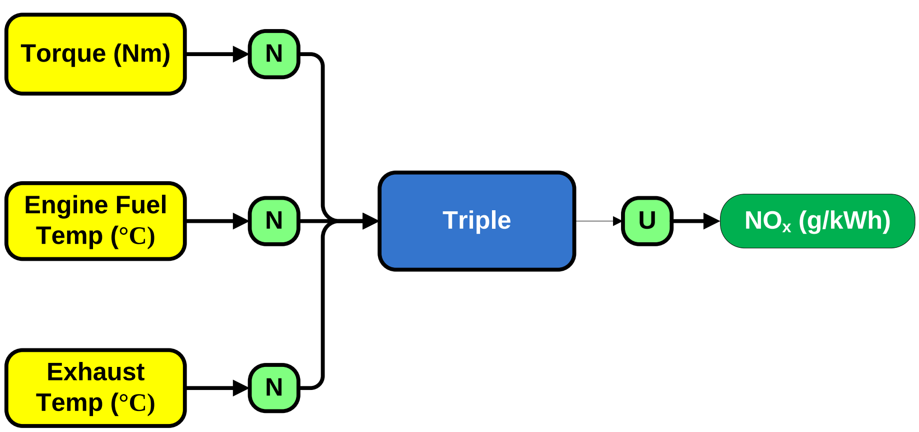

Figure 2. Polynomial network generated from ISO training data.

The EXP training data were likewise used to develop a NPN model to fit an equation to NO based on the 16 input variables. The resulting polynomial network is depicted in Figure 3. The five input variables used in the model were torque, fuel temperature, ambient temperature, engine speed, and atmospheric pressure. The predicting network accounted for approximately 96% of the observed variance in NO and consisted of the following network of equations:

$$T_n = -1.9012 + 0.0083 * T \tag{10}$$

$$FT_n = -10.4711 + 0.2144 * FT \tag{11}$$

$$AT_n = -7.7861 + 0.3021 * AT \tag{12}$$

$$ES_n = -5.6285 + 0.0028 * ES \tag{13}$$

$$AP_n = -375.0572 + 0.3762 * AP \tag{14}$$

$$\begin{aligned} TR_1 = {} & -0.4399 + 0.3631 * T_n + 0.1181 * T_n^2 - 0.149 * T_n^3 - 0.0732 * T_n * FT_n^2 \\ & + 0.1122 * AT_n + 0.2146 * T_n * AT_n - 0.1313 * T_n^2 * AT_n + 0.1614 * FT_n \\ & * AT_n - 0.2578 * T_n * AT_n^2 \end{aligned} \tag{15}$$

$$\begin{aligned} TR_2 = {} & -0.2119 + 0.2492 * TR_1 + 0.3304 * TR_1^2 + 0.1591 * TR_1^2 * ES_n + 0.1111 * \\ & TR_1 * ES_n^2 - 0.0693 * TR_1^2 * AP_n - 0.0904 * TR_1 * ES_n * AP_n \end{aligned} \tag{16}$$

$$NO = 2.8616 + 3.9914 * TR_2 \tag{17}$$

where:
T_n is normalized torque
T is observed torque (Nm)
FT_n is normalized engine fuel temperature
FT is observed engine fuel temperature (°C)
AT_n is normalized ambient temperature
AT is observed ambient temperature (°C)
ES_n is normalized engine speed
ES is observed engine speed (rpm)
AP_n is normalized atmospheric pressure
AP is observed atmospheric pressure (mbar)
TR_1 is an intermediate output of a network node with three inputs
TR_2 is an intermediate output of a network node with three inputs
NO is nitrogen oxide emissions (g/kWh).

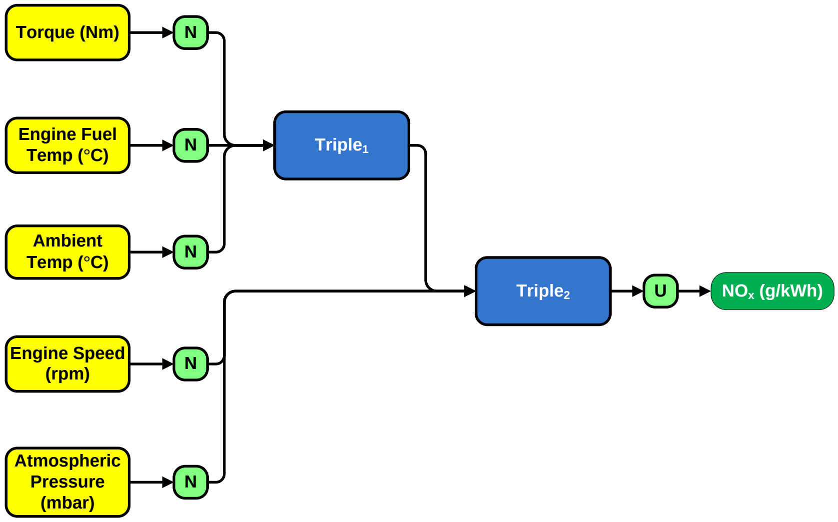

Figure 3. Polynomial network generated from EXP training data.

The NPNs of the ISO and EXP training data were applied to the evaluation data to predict NO. Table 5 summarizes the R^2, mean absolute error, and maximum absolute error of each model applied to the training data and evaluation data. When applied to the evaluation data, the R^2 for the NPN model based on the ISO training data dropped from 0.99 to 0.81. This is a considerable improvement over the LMR model (R^2 = 0.51) trained with the same data. For the NPN model applied to the evaluation data, mean absolute error and maximum absolute error increased by multiples of 5.9 and 10.9, respectively. When applied to the evaluation data, the mean absolute error of the NPN ISO based model was not different from the LMR EXP based model ($t_{(55)}$ = 0.96; p = .3411). The NPN model based on the ISO training data indicates potential to predict NO emissions at various speed and load combinations, although greater accuracy would be preferred.

The second NPN model developed with the EXP training data had an R^2 of 0.96 and maintained near the same performance with the evaluation data (R^2 = 0.95). Mean and maximum absolute errors in predicting NO for evaluation data based on the EXP model were multiples of 2.5 and 1.3, respectively, of the training data. This error was less than half the mean absolute error of the ISO based model. When applied to the evaluation data, the mean absolute error of the NPN EXP based model was significantly lower than the NPN ISO based model ($t_{(55)}$ = 5.078; $p < 0.0001$). The EXP emission tests with its additional test modes are preferred for modeling NO.

The relative accuracy of the LMR and NPN models based on the ISO data, at predicting the evaluation data, are illustrated in Figure 4. When actual NO ranged from 1.3 to 4.0 g/kWh, the LMR model predicted NO values ranging from -2.4 g/kWh to 13.8 g/kWh. In contrast, the NPN model predicted values ranged from 1.3 g/kWh to 6.9 g/kWh. Both models under predicted when actual NO was above 4.0 g/kWh and the highest actual output of nearly 36 g/kWh (low idle, no load condition) is where each model had its maximum absolute error.

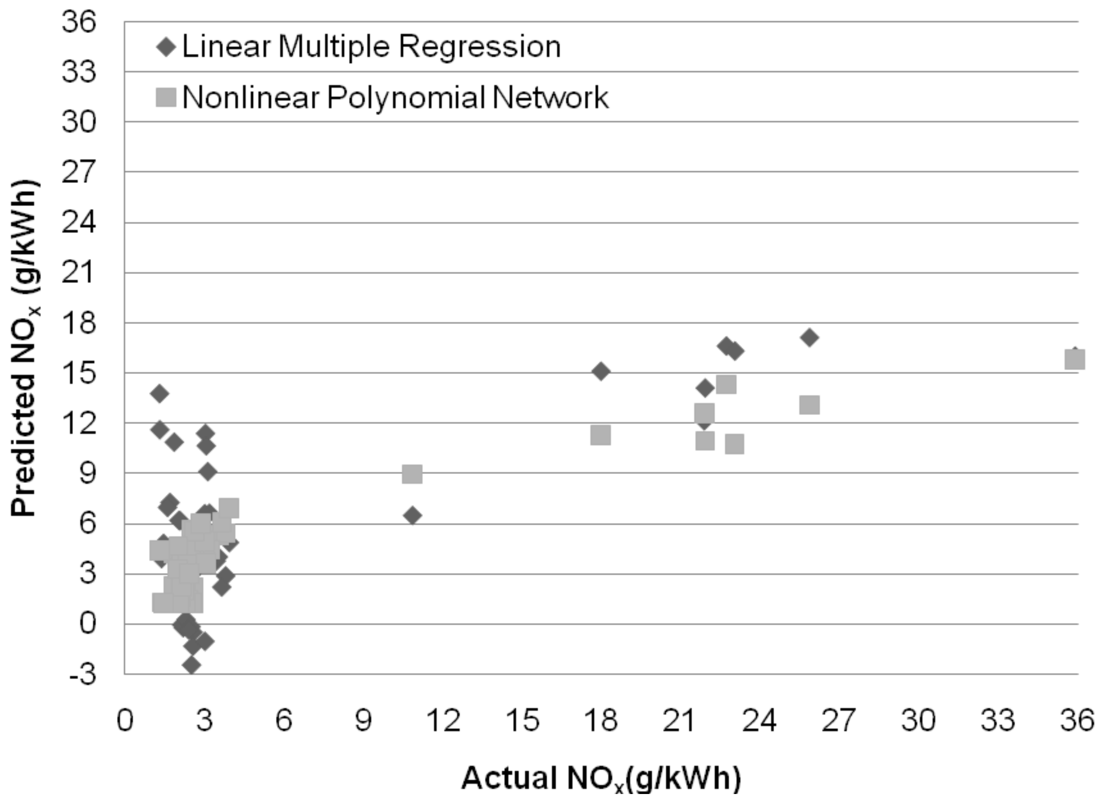

Figure 4. Predicted NO values for evaluation data, from LMR and NPN models, derived from ISO training data.

The relative accuracy of the LMR and NPN models based on the EXP data, at predicting the evaluation data, are depicted in Figure 5. When NO was actually between 1.3 and 4.0 g/kWh, the LMR model predicted NO values ranging from -0.8 g/kWh to 7.0 g/kWh. In comparison, the NPN model predicted values ranged from 1.3 g/kWh to 4.6 g/kWh. The LMR model under predicted when actual NO was above 4.0 g/kWh. When NO was above 4.0 g/kWh, the NPN predictions were closer to actual values. Maximum absolute error occurred for both models at the highest actual output of nearly 36 g/kWh, which occurred under low idle, no load conditions.

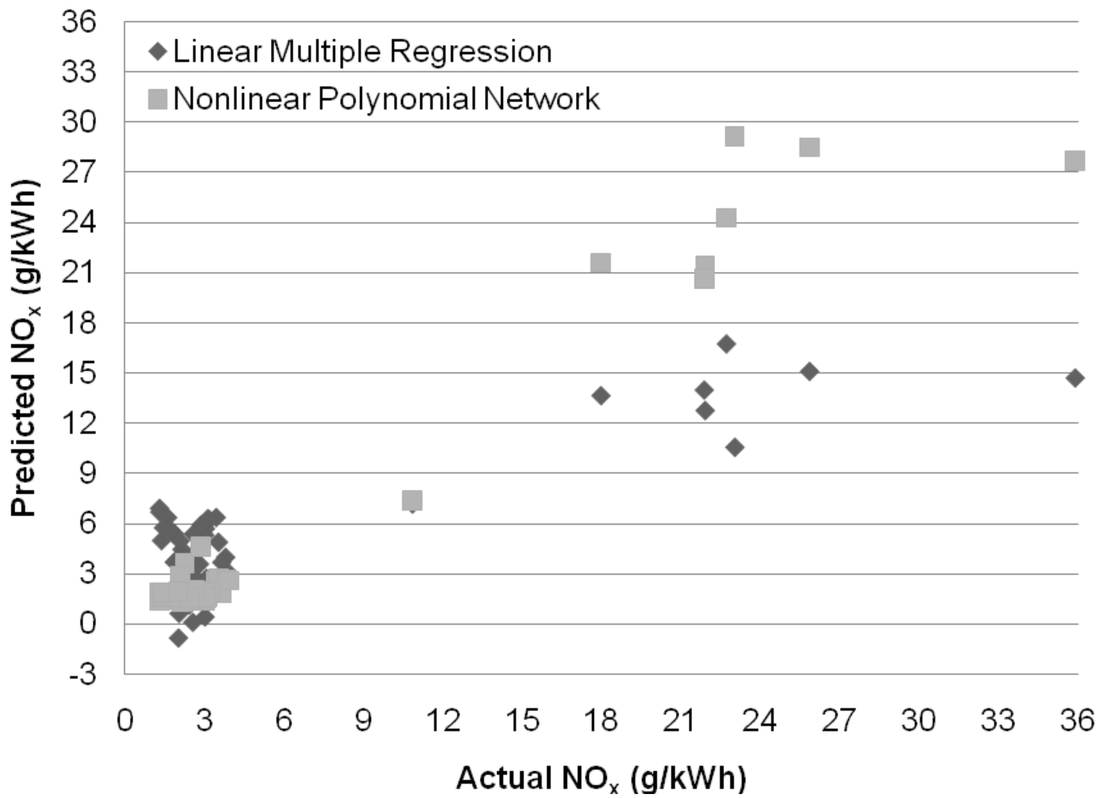

Figure 5. Predicted NO values for evaluation data, from LMR and NPN models, derived from EXP training data.

Actual NO emissions of greater than 4.0 g/kWh occurred when the engine was operating at no load. Although the maximum error for all models occurred at no load, under normal conditions, an irrigation engine would be operating under load at all times. The NPN model based on the EXP data best modeled the full range of emission values.

The Hansson et al. (2001) conclusion that it was not possible to design one set of emissions factors that produced representative results for all types of tractors and work operations, is not disputed. However, results of this study indicate that if a broader range of data were made available from ISO 8178-4 type emission tests with additional test modes, it may be possible to predict NO emissions of an engine operating at a constant load and speed using NPN. An emission test would need to be replicated three times to provide the same amount of observations used to train the models in this study.

Both of the NPN models used torque and fuel temperature as inputs. The ISO based model also included exhaust temperature, whereas the EXP based model added ambient temperature, engine speed, and atmospheric pressure. The selection of certain inputs by the NPN process does not mean other input combinations are not as effective at modeling NO. As the NPN is developed, the first combination of inputs that results in the best PSE is retained and subsequent input combinations are discarded unless PSE is improved.

Instruments are readily available to measure exhaust temperature, ambient temperature, engine speed, and atmospheric pressure. Torque data for an operating engine could be derived from the available percent torque on the CAN. Assuming resolution of the percent torque value is 1% (SAE 2002, spn513), the reduced resolution may adversely affect a model. Another option, for research purposes, would be to install strain gage transducers as used by Hansson et al. (2003).

4. SUMMARY AND CONCLUSIONS

This study of using ISO and EXP data sets with LMR and NPN modeling to predict NO produced the following results:

- LMR using the ISO training data (R^2 = 0.94) resulted in over-training of the model, as applied to the evaluation data (R^2 = 0.51).
- LMR using the EXP training data (R^2 = 0.60) resulted in a better fit for the evaluation data (R^2 = 0.73) than the ISO training data, but the model under-preformed compared to the NPN models.
- NPN using the ISO training data (R^2 = 0.99) resulted in a better fit for the evaluation data (R^2 = 0.81), than either of the LMR models.
- NPN using the EXP training data (R^2 = 0.96) resulted in the best model when applied to the evaluation data (R^2 = 0.95) and is recommended for predicting NO.
- This study suggests it may be possible to collect data during ISO 8178-4 type emission tests, with additional test modes, and model with NPN to predict NO emissions for a diesel engine operating at various constant speeds and loads.

5. REFERENCES

Abdel-Aal, R. E., Mangoud, A. M. 1997. Modeling Obesity Using Abductive Networks. *Computers and Biomedical Research* 30: 451-471.

Agarwal, A. 1999. Abductive Networks for Two-Group Classification: A Comparison with Neural Networks. *The Journal of Applied Business Research* 15: 1-12.

Barron, A. R. 1984. Predicted Squared Error: A Criterion for Automatic Model Selection. In S. J. Farlow (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms* (Vol. 54, p. 350). New York, NY: Marcel Dekker, Inc.

Barron, A. R., Barron, R. L. 1988. Statistical Learning Networks: A Unifying View. In E. J. Wegman, D. T. Gantz, J. J. Miller (Ed.), Computing Science and Statistics: Proceedings of the 20th Symposium on the Interface, 192-203. Alexandria, VA: American Statistical Association.

Barron, R. L., Mucciardi, A. N., Cook, F. J., Craig, J. N., Barron, A. R. 1984. Adaptive Learning Networks: Development and Applications in the United States of Algorithms Related to GMDH. In S. J. Farlow (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms* (Vol. 54, p. 350). New York, NY: Marcel Dekker, Inc.

Cerullo, M. J., Cerullo, M. V. 2006. Using Neural Network Software as a Forensic Accounting Tool. *Information Systems Control Journal* 2: 1-5.

Cooper, D. A., Andreasson, K. 1999. Predictive NOx emission monitoring on board a passenger ferry. *Atmospheric Environment* 33: 4637-4650.

Drake, K. C., Kim, R. Y. 1997. Abductive Information Modeling Applied to Financial Time Series Forecasting. In: Nonlinear Financial Forecasting, Finance and Technology: 95-108.

Drake, K. C., Kim, R. Y., Kim, T. Y., Johnson, O. D. 1994. Comparison of polynomial network and model-based target recognition. Proceedings for the Sensor Fusion and Aerospace Applications II: 2-11. Bellingham, WA: SPIE.

Duffy, J. J., Franklin, M. A. 1975. A Learning Identification Algorithm and Its Application to an Environmental System. *IEEE Transactions on Systems, Man, and Cybernetics* 5: 226-240.

Farlow, S. J. 1984. The GMDH Algorithm. In S. J. Farlow (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms* (Vol. 54, p. 350). New York, NY: Marcel Dekker, Inc.

Griffin, M. P., Scollan, D. F., Moorman, J. R. 1994. The Dynamic Range of Neonatal Heart Rate Variability. *Journal of Cardiovascular Electrophysiology* 5: 112-124.

Hansson, P.-A., Lindgren, M., Noren, O. 2001. A Comparison between Different Methods of calculating Average Engine Emissions for Agricultural Tractors. *Journal of Agricultural Engineering Research* 80: 37-43.

Hansson, P.-A., Lindgren, M., Nordin, M., Petersson, O. 2003. A Methodology for Measuring the Effects of Transient Loads on the Fuel Efficiency of Agricultural Tractors. *Applied Engineering in Agriculture* 19: 251-257.

Hogan, J. A., D. G. Watson, and T. V. Harrison. 2007. Data Points and Duration for Estimating Fuel Consumption of a LPG Engine. *Agricultural Engineering International* IX: 1-10.

ISO. 1996. ISO 8178-4; Reciprocating internal combustion engines, exhaust emission measurements, Part 4: test cycles for different engine applications. Genéve, Switzerland: International Organization of Standardization.

Ivakhnenko, A. G. 1968. The Group Method of Data Handling -- A Rival of the Method of Stochastic Approximation. *Soviet Automatic Control* 13: 43-55.

Ivakhnenko, A. G. 1971. Polynomial Theory of Complex Systems. *IEEE Transactions of Systems, Man, and Cybernetics* 1: 364-378.

Ivakhnenko, A. G., Stepashko, V. S., Khomovnenko, M. G., Galyamin, Y. P. 1977. Self-Organization of Dynamic Models of Growth of Agricultural Crops for Control of Irrigated Crop Rotation. *Soviet Automatic Control* 10: 23-33.

Kim, K. S. 2002. Value management and common accounting performance measures for corporations. *Expert Systems with Applications* 22: 331-336.

King, M. A., Elder, J. F., Gomolka, B., Schmidt, E., Summers, M., Toop, K. 1998. Evaluation of Fourteen Desktop Data Mining Tools. IEEE International Conference on Systems, Man, and Cybernetics. San Diego, CA.

Kleinsteuber, S., Sepehri, N. 1996. A Polynomial Network Modeling Approach to a Class of Large-Scale Hydraulic Systems. *Computers & Electrical Engineering* 22: 151-168.

Krijnsen, H. C., Bakker, R., van Kooten, W. E., Calis, H. P., Verbeek, R. P., van den Bleek, C. M. 2000. Evaluation of Fit Algorithms for NOx Emission Prediction for Efficient DeNOx Control of Transient Diesel Engine Exhaust Gas. *Industrial & Engineering Chemical Research* 39: 2992-2997.

Lebow, W. M., Mehra, R. K., Toldalagi, P. M. 1984. Forecasting Applications of GMDH in Agricultural and Meteorological Time Series. In S. J. Farlow (Ed.), *Self-Organizing Methods in ModeilngL GMDH Type Algorithms* (p. 350). New York: Marcel Dekker.

MarketMiner Inc. 2004. MarketMiner ModelQuest Analyst Version 6.0 User's Guide. Charlottesville, VA: MarketMiner Inc.

Montgomery, G. J. 1989. Abductive Diagnostics. AIAA Computers in Aerospace Conference, 7th, 267-275. Washington, DC: American Institute of Aeronautics and Astronautics.

Montgomery, G. J., Drake, K. C. 1991. Abductive reasoning networks. *Neurocomputing* 2: 97-104.

Montgomery, G. J., Hess, P., Hwang, J. S. 1990. Abductive Networks Applied to Electronic Combat. *Applications of Artificial Neural Networks* 1294: 454-465. Bellingham, WA: SPIE.

National Instruments. 2003. LabVIEW 7 Express User Manual. Austin, TX: National Instruments.

Pachepsky, Y. A., Rawls, W. J. 1999. Accuracy and Reliability of Pedotransfer Functions as Affected by Grouping Soils. *Soil Science Society of America Journal* 63: 1748-1757.

Pyo, S., Uysal, M., Chang, H. 2002. Knowledge Discovery in Database for Tourist Destinations. *Journal of Travel Research* 40: 396-403.

Ramamurthy, R., Clark, N. N., Atkinson, C. M., Lyons, D. W. 1998. Models for Predicting Transient Heavy Duty Vehicle Emissions. SAE Paper 982652. Warrendale, PA: SAE.

Reddy, V. R., Pachepsky, Y. A. 2000. Predicting crop yields under climate change conditions from monthly GCM weather projections. *Environmental Modeling & Software* 15: 79-86.

SAE. 2002. SAE J1939-71 AUG2002 Recommended Practice for Truck and Bus Control and Communications Network—Vehicle Application Layer. Warrendale, PA: SAE.

SAS. 2007. SAS 9.1. Cary, NC: SAS Institute, Inc.

Silis, Y. Y., Rozenblit, A. B. 1976. Algorithm for Construction of Decision Function in the Form of a Complex Logic Proposition. *Soviet Automatic Control* 9: 1-5.

Stepanov, V. A. 1974. Results of Mass Checking of GMDH Efficiency in Long-Term Prediction of Demand for Consumer Goods. *Soviet Automatic Control* 7: 46-51.

Watson, D. G., T. V. Harrison, and R. W. Steffen. 2008. Data Points and Duration for Estimating Fuel Consumption of a Diesel Engine. *Agricultural Engineering International* X: 1-9.

Yanowitz, J., Graboski, M. S., McCormick, R. L. 2002. Prediction of In-Use Emissions of Heavy-Duty Diesel Vehicles from Engine Testing. *Environmental Science & Technology* 36: 270-275.